# Correlation of Bone Quality to Biomechanical Response: Large-Scale Study Methodology

Undergraduate Honors Thesis

Presented in Partial Fulfillment of the Requirements for the Degree Bachelor of Science with Research Distinction in the School of Health and Rehabilitation Science of The Ohio State University

Amie J. Draper

Undergraduate Biomedical Science Major

School of Health and Rehabilitation Science

The Ohio State University

April 2013

Defense Committee:

Dr. Amanda M. Agnew, Project Advisor

Dr. John H. Bolte IV, Committee Member

Dr. D. Michele Basso, Committee Member

Abstract

It has been previously shown that the quality of human bone can affect its fragility. This can have important implications in the field of injury biomechanics, in which attempts are made to elucidate the conditions under which fractures may occur. To understand the results of such research, we must establish what constitutes a "normal" bone quality and how bone quality relates to biomechanical response. Prior work has shown that bone microstructure of bilateral rib pairs varies significantly between individuals but does not vary within an individual. This implies that bilateral rib pairs can be used to examine the relationship between bone quality and biomechanical response. Previously employed methods, however, are inefficient to utilize in a large population study and an efficient, standardized methodology to assess the overall quality of a bone and relate it to its biomechanical response is needed. The purpose of this work is to develop improved methodologies in order to lay the groundwork for the development of a large-sample study of the relationship between bone quality and biomechanical response. Forty cross-sectional slides of bilateral rib pairs from 10 subjects for which variable measurements are available were utilized as the initial dataset. Methods to improve construction of cross-section images, facilitate microfeature detection and standardize statistical methodology were developed and evaluated. This study defined new methodologies to automate the image reconstruction process and developed a new, standardized statistical model that can be utilized in a large-sample study. Methods to facilitate microfeature detection in the rib cross-section images were not fully developed for use in a large-sample study, but this study laid the initial groundwork for such methods. These methodological improvements can be built into the framework for an efficient,

standardized methodology to investigate the relationship between bone quality and biomechanical response in a normal population of human ribs.

Acknowledgments

I would like to take this opportunity to thank Dr. Amanda M. Agnew for her unwavering support and guidance throughout the duration of this project. Under her direction I was afforded the opportunity to strengthen my understanding of anatomy and statistics, in addition to being exposed to significant experience in research, which is an opportunity missed by many undergraduates. Additionally, I would like to thank Dr. John H. Bolte IV for agreeing to take me on as an undergraduate student in his lab for the past two and a half years, despite my atypical background in comparison to the engineering students who primarily work in the lab. I have gained invaluable knowledge and the experience has helped me discover more about the future career path I have chosen to pursue. I would also like to acknowledge Dr. Siddharth Samsi, who developed the automated feature detection system utilized in this study, and Zhiguang Xu, the statistical consultant on this project. Finally, I would like to thank all of the members of the Injury Biomechanics Research Center, who have made my experience in the lab for the past two and a half years both enjoyable and memorable.

Table of Contents

List of Tables

List of Figures

## Chapter 1: Introduction

Previous research has shown that the quality of human bone can affect its fragility (Ebacher et al., 2007; Frost, 2002; Turner, 2002). The microstructure of bone as well as the distribution and amount of bone both contribute to its overall quality (Frost, 2002) and are important features to consider when evaluating the biomechanical response of a specific body region. The biomechanical response of a body region is particularly important in the field of injury biomechanics, in which attempts are made to elucidate the conditions under which a specific type of injury, such as a fracture, will occur. When determining the conditions under which a fracture occurs in a "typical" human, it is first necessary to establish what the "normal" overall quality of a bone is and how it can influence biomechanical response.

Some studies have characterized the naturally occurring microstructure of bones (Agnew, 2011; Rose, 2011), but published research has not yet correlated the findings to the biomechanical response of the bone in a large-scale study. Research on elderly human ribs conducted by Agnew (2011) has laid the initial groundwork for a large-scale study of the relationship between bone quality and its corresponding biomechanical response in a population of normal human bilateral rib pairs. The research by Agnew (2011) identified and quantified four important microfeatures and three important relationships of these features that could have an impact on the biomechanical response of the rib. A discussion of the microfeatures and the important relationships observed among microfeatures can be found in Chapter 2. Methodologies for analyzing bone microstructure as introduced by Agnew (2011) are currently not feasible for use in a large sample study due to time-constraints and non-standardized methodologies. This study aims to improve upon the methodologies introduced in Agnew (2011) in order to lay the

foundation for the large-sample analysis of bone microstructure. The problems with utilizing the current methodologies as outlined in Agnew (2011) and the proposed methodological improvements are discussed in Sections 1.2 and 1.3.

### 1.1 Rationale and Significance

The development of an efficient, standardized methodology is necessary before any large sample analysis can be feasibly undertaken. To understand bone microstructure in a normal population, a large sample analysis is necessary. Understanding normal bone microstructure helps in defining normal bone quality and behavior, both of which play important roles in understanding bone diseases such as osteoporosis and in understanding the relationship between bone quality and biomechanical response. Understanding the relationship between bone quality and biomechanical response also requires a large sample study, the results of which would contribute to determining how and why rib fractures occur. This knowledge has a wide variety of applications, such as aiding in interpreting the results of cadaver and anthropomorphic test device (ATD) studies and understanding rib fractures in elderly patients and patients with bone diseases. This study takes the first steps towards designing a large-sample study to analyze the relationship between bone quality and biomechanical response in a population of normal human ribs by laying the initial methodological groundwork for the large-sample analysis of bone microstructure in human ribs.

## 1.2 Problems with Current Methodologies

Utilizing the current methods as outlined in Agnew (2011) to analyze the microstructure of rib cross-sections would not be feasible in a large-sample due to time constraints and non-standardized methodology. The staining and slide preparation processes were complicated and labor intensive, including multiple lengthy solution immersions, careful temperature monitoring, and manual mounting, cutting and grinding of slide sections. After slide preparation a large amount of time was spent manually photographing sections of slides with partial overlap followed by manually stitching the photographs back together using the overlapping regions in order to obtain an image of the entire cross-section of rib with 680 pixels/nm resolution (the resolution necessary to accurately detect features). Features were then manually identified and measured in each image by a trained expert. Using these processes, it could take a few weeks to obtain data from one slide for analysis, which is not feasible for a large-sample study. Agnew (2011) analyzed data using multi-step, non-standardized statistical methodologies involving SAS, Minitab and hand calculations. The statistical methods were tailored specifically to the dataset under study and may not be easily applied to independent, larger datasets.

## 1.3 Proposed Methodological Improvements and Hypotheses

The main goal of this work was to identify ways to improve the methodologies initially proposed by Agnew (2011) in order to design an efficient method for analyzing bone microstructure in a large sample of human ribs. Staining and slide preparation processes as outlined in Agnew (2011) cannot be altered without losing the ability to detect microfeatures. Therefore, the main objectives of this study were to 1) identify and evaluate methods to improve

the efficiency of obtaining rib cross-section images, 2) identify and evaluate methods to facilitate microfeature detection and measurement and 3) to develop a standardized statistical model for use in a large sample. The proposed improvements to the methodology by Agnew (2011) are discussed here and the evaluations of the practicality of each proposal are discussed in the methods, results and discussion sections of this document.

### 1.3.1 Image Obtainment Methods

At the start of this study, machines that would allow full slide image scans at a 680 pixels/nm resolution were not readily available for slides with samples as thick and as dense as the rib cross-sections. It was concluded that until such machines became accessible, image sections would have to be manually photographed with regions of partial overlap and reconstructed into a complete rib cross-section image. Software programs capable of reconstructing images using regions of partial overlap were investigated as alternatives to manual reconstruction of images. Microsoft Research Image Composite Editor (ICE) and Autopana Giga 2.6 by Kolor are both advanced panoramic image stitching software programs and were selected for evaluation in this study. It was hypothesized that both software programs would be capable of producing composite images comparable to the manual reconstructions.

### 1.3.2 Feature Detection Methods

Manual identification and measurement of rib microfeatures is a labor-intensive process which can require a significant amount of time when performed by experts, who are few in number due to the years of experience and training needed to be qualified in this specialty. This

process could be facilitated by using either an automated microfeature detection system, which would require time but less labor, or by using novices to identify and measure microfeatures, which is a labor-intensive option, but can save time since more novices would be available than experts. Initial alternative microfeature detection methods were explored for porosity, and future investigations will include looking for alternative methods for identifying and measuring microcracks within the rib cross-sections as well. An automated system utilizing MATLAB to identify and measure pores was developed in conjunction with the Ohio Supercomputer Center, which at its most basic level utilizes color contrasts, textures and patterns in the rib cross-section images to identify and measure pores. The novice proposed in this study is a true novice, having received no prior training or information on identifying pores in the rib cross-section images. For an alternative microfeature detection method to be viable, it must exhibit a level of accuracy in both microfeature identification and measurement. For porosity, it was hypothesized that both the novice and the automated system would exhibit high identification accuracy and no significant difference in measurements when compared to the expert.

### 1.3.3 Statistical Models

The statistical methods employed in Agnew (2011) (referred to from now on as the "original model") involved multiple steps and were tailored specifically to the dataset in the study in order to deal with both the small sample size and the inter-dependency of the data. In the original model, differences in subject and rib side variation were assessed with two-way, mixed-model ANOVAs performed on data from the cutaneous and pleural cortices separately. Separate evaluations of the cortices allowed for analysis of the small sample size without sacrificing

statistical power by increasing the degrees of freedom for the datasets. Hand calculations were performed to adjust the ANOVA results for the inter-dependency of the data. Cortex differences were assessed by averaging the values of the microfeatures for the four slides for each subject and performing a paired t-test on the averaged values of each cortex. With a larger sample size, degrees of freedom become larger and less of a cause for concern in analyses, allowing for development of a more standardized statistical model. In conjunction with the Ohio State University Statistical Consulting Service, two SAS models were proposed for use with large sample sizes that accommodate for the inter-dependency of the data using nested, mixed-model ANOVAs. Model 1 analyzes the combined pleural and cutaneous data in one nested, mixed-model ANOVA. Model 2 is similar to the original model in that it analyzes the pleural and cutaneous data in two separate nested, mixed-model ANOVAs and the results are interpreted together. Model 1 is the ideal model for a large-scale study as it is simpler to use and produces results that are easier to interpret, however, model 2 was proposed to determine whether or not splitting the data into datasets by cortex substantially changes the microfeature relationships observed.

If either of the newly proposed models is to be utilized in future research, it should produce similar results to those obtained in the original model for percent porosity, crack length, crack density and crack surface density. While it is not expected that the p-values from the models will match due to differences in types of analyses performed, it is expected that subject, rib side, and cortex variation will exhibit similar trends within the same dataset. Expected trends based on the original model for subject, rib side and cortex variation are summarized for each

microfeature in Table 1. It was hypothesized that model 2 would be more likely to exhibit trends closer to the original model than model 1 due to its similar design.

| Microfeature | Expected Subject Variation | Expected Rib Side Variation | Expected Cortex Variation |
|---|---|---|---|
| Percent Porosity | Significant variation | No significant variation | Significant variation |
| Crack Length | Signficican variation | No significant variation | No significant variation |
| Crack Density | Significant variation | No significant variation | No significant variation |
| Crack Surface Density | Significant variation | No significant variation | No significant variation |

**Table 1. Expected trends in subject, cortex and rib side variation for microfeatures.** These expectations are based on the original model.

If either model produces similar results to the original model, the new model will be used to explore additional questions that could not be addressed using the original model.

The original study measured each microfeature by quadrants in each cross-sectional image, defined by areas seen in Figure 1. Agnew (2011) explored the differences in the cutaneous and pleural cortices for the microfeatures, however superior and inferior side differences in each cortex were not explored due to the complexity such an analysis would add to the statistical model. If significant variation between the superior and inferior quadrants of each half of the cortex exists, it could provide additional information regarding the mechanical loading patterns to which ribs are adapted and sides of cortices should be considered as an additional variable in future analyses of rib microstructure. It was hypothesized that no

significant variation between superior and inferior sides within the cutaneous and pleural cortices would exist.

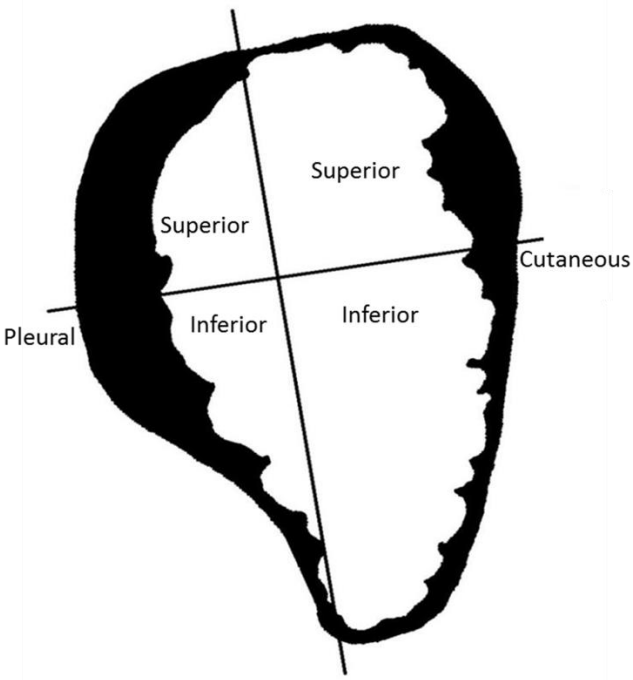


**Figure 1. Graphical definition of rib quadrants.** Data were collected for ribs in quadrants. Halves were defined by the pleural and cutaneous cortices, quadrants defined by the inferior and superior sides of each cortex. This silhouette is oriented in the anatomical position of the rib.

The original model proposed in Agnew (2011) was designed for the analysis of 6$^{th}$ bilateral rib pairs and could not be easily adjusted to accommodate samples with additional rib levels. Both models 1 and 2 were developed to be easily adjusted to accommodate various rib levels in a subject. If a significant variation in rib levels exists, it would suggest that bilateral rib pairs from the same thoracic level should be utilized in a large sample study of the relationship of bone quality to biomechanical response. It was hypothesized for an independent dataset involving multiple rib levels that rib levels would exhibit no significant variation between one

another and trends for subject variation, cortex variation, and rib side variation would be the same as those observed in the dataset including only one rib level.

# Chapter 2: Microfeatures and Relationships

## 2.1 Microfeatures

The microfeatures discussed in this section are quantified values of two naturally occurring phenomena within cortical bone: cortical porosity and microdamage (also referred to as microcracks or microfractures). Cortical porosity refers to the volume fraction of cortical bone that is not occupied by bone tissue (Laval-Jeantet et al., 1983), and microcracks refer to the very small fractures that can occur as a result from fatigue in bone (Taylor, 1997). Several studies have shown that both cortical porosity and microcracks can impact the mechanical properties of bone, such as strength and stiffness (Burr et al., 1997; Burr et al., 1998; Ebacher et al., 2007; Sobelman et al., 2004; Zioupos, 2001), which suggests that variation in these properties between individuals may explain some observed differences in biomechanical response. Agnew (2011) quantified one microfeature relating to cortical porosity, termed percent porosity, and three microfeatures relating to microcrack damage, termed crack length, crack density, and crack surface density.

Percent porosity is defined as the area of intracortical porosity relative to the defined cortical area and has no units. In a mathematical equation it can be written as:

1.) Percent Porosity = (Total cortical porosity area)/(Total cortical bone area)

Crack length is defined as the mean length of all the microcracks, expressed in μm. Mathematically it can be written as:

2.) Crack Length = (Total summation of microcracks lengths)/(Total number of microcracks)

Crack density is defined as the number of microcracks relative to the true amount of cortical bone area, expressed in $mm^2$. In a mathematical equation crack density is written as:

3.) Crack Density = (Total number of microcracks)/(Total cortical area – Total cortical porosity area)

Crack surface density is defined as the total area of microcracks relative to the true amount of cortical bone area, expressed in $\mu m/mm^2$. Mathematically, crack surface density can be written as:

4.) Crack Surface Density = (Number of microcracks)*(Crack length)/(Total cortical area – Total cortical porosity area)

All of these quantified microfeatures provide information about microcracks and porosity within the rib of an individual and have the potential to be correlated with data from mechanical tests in future studies of bilateral rib pairs. Such correlations will aid in understanding the relationship between bone quality and biomechanical response.

## 2.2 Important Microfeature Relationships

Agnew (2011) identified three relationships among the microfeatures that have important implications for future research in bone quality and biomechanical response. These three relationships are briefly discussed here and are given new names for ease of discussion through the remainder of the paper.

*Relationship 1: All features showed significant inter-individual variation.* This relationship implies that differences in individual bone microstructure could explain some of the

observed individual differences in fragility in mechanical tests. From this point on, the microfeature relationships observed between individuals will be referred to as subject variation.

*Relationship 2: All features showed no significant variation within the bilateral ribs of an individual*. In simpler terms, this relationship states that within an individual, there is no significant difference in microfeatures between left and right ribs of the same thoracic level. This relationship implies that future studies can utilize bilateral rib pairs to examine the relationship between bone microstructure and biomechanical response—one rib can be used for pre-test microstructure analysis and one rib can be used for mechanical testing (since both processes are destructive). For the remainder of this document, the microfeature relationships observed between bilateral rib pairs within an individual will be referred to as rib side variation.

*Relationship 3: Some features showed significant variation between the pleural (compressive) and cutaneous (tensile) cortices.* Mechanical tests of ribs utilize different loading patterns, which can have effects on the different cortices of the rib. For example, during bending the cutaneous cortex of the rib experiences tension while the pleural cortex of the rib experiences compression. Differences in the microfeatures of the cortices could contribute to understanding how well ribs are adapted to certain mechanical loading patterns, which can aid in interpreting results of mechanical tests such as bending. For the remainder of this document, the microfeature relationships observed between the pleural and cutaneous cortices will be referred to as cortex variation.

# Chapter 3: Materials and Methods

## 3.1 Sample and Data Collection

Cross-sections of the right and left 6th ribs were obtained from 10 fresh elderly post-mortem human subjects (PMHS). International Review Board approval was not required for these samples as cadavers are not considered human subjects for review purposes. The elderly PMHS were comprised of 5 males and 5 females, all of whom were over the age of 70 at time of death and who had experienced no recent thoracic trauma. Two cross-sections of each rib, taken at 50% of the total rib length, were obtained for a total of 40 cross-sections. Each cross-section is 60 μm thick, mounted to a standard microscope slide and stained with Basic Fuchsin Hydrochloride. The complete methodology for preparation of the slides has been described in detail in previous research (Agnew, 2011). Percent porosity, crack length, crack density and crack surface density had been previously quantified for all 40 slides, and those measurements comprised the large data-set utilized in the methodological developments of this study.

Right and left $5^{th}$, $6^{th}$ and $7^{th}$ ribs were obtained at autopsy from 12 pediatric post-mortem human subjects following full review and approval by an Institutional Review Board. Ribs were subjected to three-point bending tests in which the cutaneous cortex experienced compressive loads and the pleural cortex experienced tension. Following experimental testing, rib cross-sections were obtained from 1 cm near the location of impact. Each section is 50 μm thick, mounted to a standard microscope slide and photographed under a linear polarized light filter. The complete methodologies for the experimental tests, slide preparations and photography have been described in detail in previous research (Agnew et al., in press). The sample utilized in this study included 48 cross-sections comprised of $5^{th}$, $6^{th}$ and/or $7^{th}$ bilateral rib pairs from 8

subjects. Percent porosity was previously quantified for the images and was utilized as an independent dataset for analysis of rib level variation.

### 3.2 Image Obtainment Methods

Forty images of rib-cross sections were reconstructed by Microsoft Research ICE and Autopana Giga 2.6 and the reconstructions were compared back to the manually stitched image cross-sections. Reconstructions were only considered an exact match if there was no observable difference between the automated reconstruction and the manual reconstruction (Figure 2). The percentage of the 40 images correctly reconstructed was used to evaluate the accuracy of each software program. Image editing capabilities of each software program were also investigated.

**Figure 2. Automated image reconstruction comparisons.** Three image reconstructions for the same image were obtained using Microsoft Research ICE (left), manual reconstruction (middle) and Autopana Giga 2.6 (right). For this image, the leftmost image would be considered an accurate reconstruction (i.e. an exact match) while the rightmost image would be considered an inaccurate reconstruction.

### 3.3 Feature Detection Methods

An expert, a novice and the automated system (termed here as "observers") each identified and measured porosity on 10 images. The 10 images consisted of the first section of the left rib from each subject in order to better examine the capabilities of the observers on a heterogeneous set of bone microstructures. Each observer was evaluated for accuracy in pore identification and accuracy in pore area measurement. The expert was treated as the "true" set of pores when assessing pore identification accuracy. Pore identification accuracy was evaluated using two different percentages:

1.) Percent of Correct Pores = [(Number of pores identified by novice or automated system also identified by the expert)/(Total number of pores identified by expert)] X 100%

2.) Percent of False Pores = [(Number of pores identified by novice or automated system that were not identified by the expert)/(Total number of pores identified by novice or automated system)] X 100%

For pore area measurement accuracy, significant variations in area measurements between the possible pairs of observers were identified. Only pores that were correctly identified by both observers in a pair were included in the assessment. Data was not normally distributed after performing various transformations, therefore significant observer differences were identified using sign tests on the untransformed differences of area measurements in a pair of observers. P-values $<0.05$ were considered significant variations in measurements between observers.

### 3.4 Statistical Models

For all analyses, p-values $<0.05$ were considered significant. Model 1 and model 2 were utilized to re-analyze the microfeatures in the original dataset. The results of each model for subject variation, rib variation and cortex variation were compared to the results from the original model and differences were identified. Significant variation in the superior and inferior sides of the cutaneous and pleural cortices was assessed using both models. One model was selected to assess rib level variation in the pediatric dataset. Additionally, rib side variation, subject variation and cortex variation were all examined in the dataset.

# Chapter 4: Results

## 4.1 Image Reconstruction Methods

For the 40 reconstructed images, Microsoft Research ICE and Autopana Giga 2.6 exhibited 90% and 65.7% accuracy, respectively. Microsoft Research ICE exhibited no image editing capabilities, while Autopana Giga 2.6 was found to contain editing tools to manually edit an automatically reconstructed image.

## 4.2 Feature Detection Methods

Figures 3 and 4 display the results of the pore identification accuracy assessment. Figure 3 shows that the novice consistently outperformed the automated system in percentages of correct pores for each subject, with most percentages being above 80% or 90%. The automated system typically exhibited percentages of correct pores in the 70%-80% range, with the exception of subjects 1, 2 and 6. Figure 4 shows a very low percentage of false pores for the novice and a very high percentage of false pores for the automated system, with most images having 80% or higher of all pores identified as being false pores.

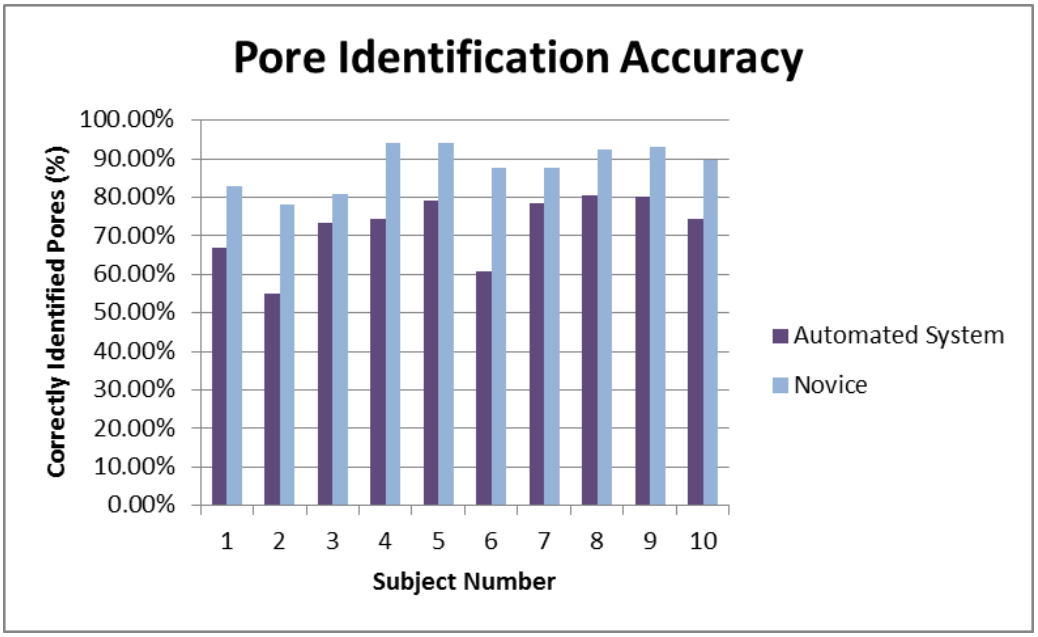


**Figure 3. Pore identification accuracies of the novice and automated system.**

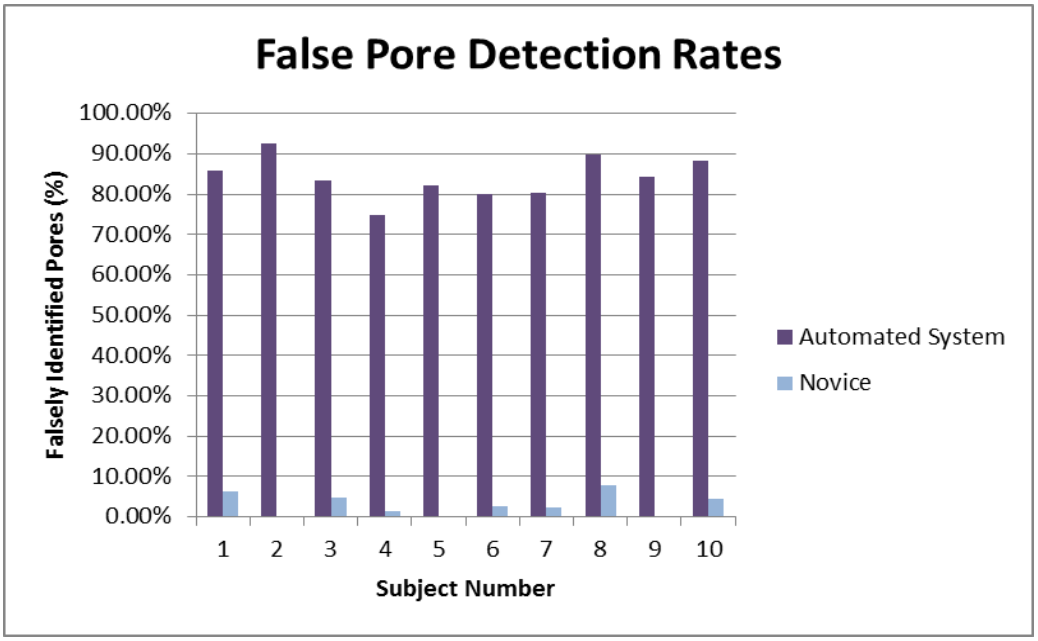


**Figure 4. False pore detection rates for the novice and the automated system.**

Table 2 gives the means and p-values obtained from the sign tests for each of the three observer comparisons on the 10 images. Most of the comparisons revealed significant variation

in measurements between observers, however a few images revealed no significant variation. The automated system showed no significant variation in measurements from the expert for images from subjects 1 and 2, and showed no significant variation in measurements from the novice for images from subjects 2, 6 and 7. The novice showed no significant variation in measurements from the expert for images from subjects 2, 5 and 7.

| | Difference Calculation | Subject # | Mean Difference | P-Value |
|---|---|---|---|---|
| **Expert Vs. Automated System** | Automated System-Expert | 1 | -1175.14 | 0.3531 |
| | | 2 | -442.273 | 0.0681 |
| | | 3 | -5324.85 | **0.0004** |
| | | 4 | -757.57 | **0.0064** |
| | | 5 | -2807.30 | **<0.001** |
| | | 6 | -761.547 | **0.002** |
| | | 7 | -691.594 | **0.0009** |
| | | 8 | 81.5211 | **<0.0001** |
| | | 9 | -576.768 | **0.0005** |
| | | 10 | -250.2241 | **<0.0001** |
| **Expert Vs. Novice** | Novice-Expert | 1 | -6676.51 | **0.0368** |
| | | 2 | 746.141 | 0.4168 |
| | | 3 | -2779.99 | **0.0089** |
| | | 4 | -1924.04 | **<0.0001** |
| | | 5 | -98.7249 | 0.0922 |
| | | 6 | -1463.98 | **0.0401** |
| | | 7 | -214.3 | 0.826 |
| | | 8 | 4.3985 | **<0.0001** |
| | | 9 | -910.732 | **<0.0001** |
| | | 10 | 216.1419 | **<0.0001** |
| **Novice Vs. Automated System** | Automated System-Novice | 1 | -568.411 | **<0.0001** |
| | | 2 | -883.69 | 0.0925 |
| | | 3 | -5691.44 | **0.0035** |
| | | 4 | 150.64 | **<0.0001** |
| | | 5 | -2814.51 | **<0.0001** |
| | | 6 | -170.758 | 0.1326 |
| | | 7 | -377.41 | 0.1909 |
| | | 8 | 112.123 | **<0.0001** |
| | | 9 | 116.3464 | **<0.0001** |
| | | 10 | -412.571 | **0.0076** |

**Table 1. Inter-observer sign test results.** The difference calculation in this chart indicates how the measurement difference values were calculated for the sign tests. Significant p-values are bolded.

### 4.3 Statistical Models

Table 3 displays the p-values from model 1, model 2 and the original model for subject variation, rib side variation and cortex variation in each of the four microfeatures. Model 1 differed from the original model in cortex variation significance for percent porosity and crack surface density and in subject variation significance for crack length. For the remaining microfeatures, model 1 showed the same trends in significant variation as the original model as was expected. In the pleural cortex, model 2 differed from the original model in rib side variation significance for percent porosity and differed in subject variation significance for crack density and crack surface density. Both cutaneous and pleural cortices differed in subject variation significance from the original model. Model 2 was unable to be adapted to analyze cortex variation in this study.

| | | Model 1 | Model 2 | | Original Model | |
|---|---|---|---|---|---|---|
| | | | Cutaneous | Pleural | Cutaneous | Pleural |
| Percent Porosity | Subject Variation | **0.0277** | **0.037** | **0.0344** | **0.0001** | **0.0001** |
| | Rib Variation | 0.3465 | 0.0988 | **0.0442** | 0.965 | 0.616 |
| | Cortex Variation | 0.0888 | ---------- | | **0.04** | |
| Crack Length | Subject Variation | 0.2139 | 0.1815 | 0.4125 | **0.001** | **0.014** |
| | Rib Variation | *N/A | 0.4675 | 0.1953 | 0.676 | 0.909 |
| | Cortex Variation | 0.578 | ---------- | | 0.879 | |
| Crack Density | Subject Variation | **0.0463** | **0.0478** | 0.0792 | **0.018** | **0.036** |
| | Rib Variation | 0.2206 | 0.2251 | 0.2072 | 0.753 | 0.4404 |
| | Cortex Variation | 0.2249 | ---------- | | 0.186 | |
| Crack Surface Density | Subject Variation | **0.0357** | **0.0472** | 0.0762 | **0.015** | **0.015** |
| | Rib Variation | 0.3305 | 0.1641 | 0.1727 | 0.849 | 0.482 |
| | Cortex Variation | **0.0379** | ---------- | | 0.157 | |

**Table 2. P-values from model 1, model 2 and the original model for microfeature variation trends.** Significant p-values are bolded, and differences from the original model for Model 1 are highlighted in red and differences for Model 2 are highlighted in blue. The *N/A indicates that a p-value was not obtained from the SAS results. This occurs when a variable explains none of the variation within the dataset. Therefore, N/A is treated as an insignificant finding.

Both model 1 and model 2 were used to assess variation in the superior and inferior sides of the cutaneous and pleural cortices, the p-values of which are shown in Table 4. Neither model revealed significant variation between the sides within the cortices.

| | | Crack Density | Crack Surface Density | Crack Length | Percent Porosity |
|---|---|---|---|---|---|
| Model 1 | | 0.7675 | 0.5086 | 0.4900 | 0.1235 |
| Model 2 | Pleural | 0.8839 | 0.8957 | 0.9872 | 0.0785 |
| | Cutaneous | 0.7516 | 0.5864 | 0.2326 | 0.8774 |

**Table 3. P-values from models 1 and 2 for superior versus inferior cortex variation.**

Model 1 was selected for the analysis of the pediatric dataset because of its ability to assess cortex variation and its general agreement with the original model in variation trends. Table 5 shows the p-values from model 1 for percent porosity in the pediatric dataset. Rib level variation, rib side variation and subject variation were not significant, but cortex variation was significant.

| Variable | P-Value |
|---|---|
| Subject | 0.3696 |
| Rib Level | 0.4616 |
| Rib Side | 0.4142 |
| Surface | 0.0013 |

**Table 4. P-values from model 1 for the pediatric dataset.**

# Chapter 5: Discussion

Current published research has explored naturally occurring microstructure in bone and has also determined that inducing microdamage prior to mechanical testing can significantly influence bone response during testing (Ebacher et al., 2007). However, limited work has been done in examining the relationship between naturally occurring microstructure and biomechanical response. The main objective of this study was to improve the methodologies for analyzing bone microstructure as introduced in Agnew (2011) in order to lay the foundations upon which a large-sample study of the relationship between bone quality and biomechanical response in human ribs can be developed. Methodological improvements explored in this study included automating methods of obtaining rib cross-section images, facilitating microfeature detection and simplifying and standardizing statistical procedures.

## 5.1 Image Obtainment Methods

At the beginning of this study, machines that would allow a full rib cross-section to be scanned at 680 pixel/nm resolution were not readily available for use. Therefore, two panoramic image stitching software programs, Microsoft Research ICE and Autopana Giga 2.6, were explored as alternative options to manually stitching composite rib cross-section images. It was hypothesized that both programs would be able to reconstruct image composites comparable to the manual reconstructions. With 90% of images correctly reconstructed, the hypothesis was accepted for Microsoft Research ICE, but rejected for Autopana Giga 2.6, which only correctly reconstructed 67.5% of images. However, Autopana Giga 2.6 was found to have the capability to edit images after they had been automatically stitched together, allowing a user to stitch together

an image and then edit the incorrect portions of the image. This editing feature is not present in Microsoft Research ICE and will be useful when Microsoft Research ICE cannot correctly restitch a rib cross-section image. Since the evaluation of these software programs, a microscope that is capable of scanning full rib cross-section slides has been made available for use. Utilizing the microscope to obtain full rib cross-sections is recommended for future research, but both image software programs can be utilized in future research for partial reconstructions of rib cross-section images.

### 5.2 Feature Detection Methods

A novice, an expert and an automated system were all compared for inter-observer error in detecting and measuring porosity in the rib cross-section images. It was hypothesized that both the novice and the automated system would exhibit high identification accuracy with no significant difference in measurement when compared to the expert. Based on the novice's high percentage of correct pores identified and the low percentage of false pores identified, the novice was determined to have high identification accuracy for porosity, as was hypothesized. The automated system exhibited a fairly high percentage of correctly identified pores but exhibited a high percentage of false pores identified, which led to overall poor identification accuracy for porosity, which was unexpected.

Each possible pair of observers (expert/novice, expert/automated system, novice/automated system) exhibited significant variation from one another in pore measurements for a majority of the 10 images examined in the analysis which was not expected. Upon examination of the histogram of differences in the observer values, it was discovered that with

the exception of a few extreme outliers, many data distributions were centered close to zero, as shown in the example in Figure 5.

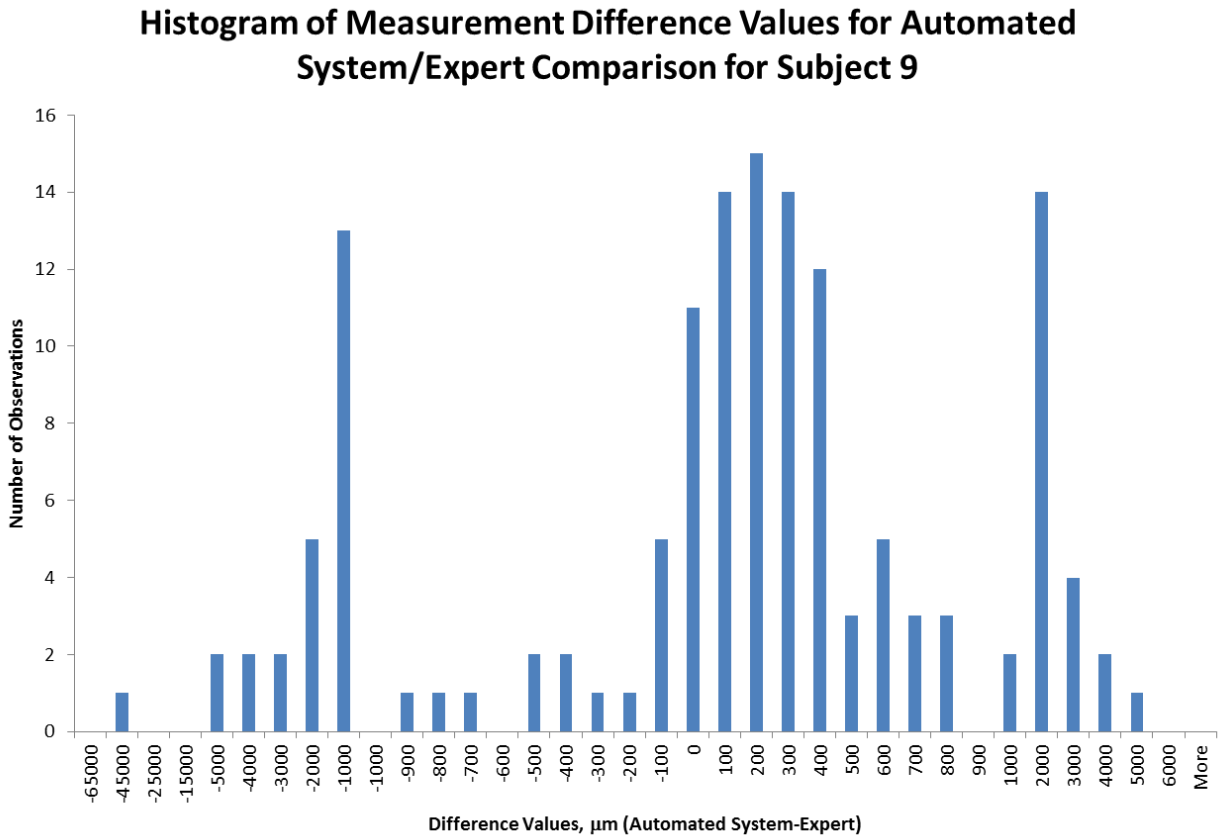


**Figure 5. Histogram of porosity measurement difference values.** Most difference values are within close range to zero.

This suggests that extreme outliers could have a significant impact on the results of the significance tests and that for a majority of the measurements the values do not differ significantly between observers. Extreme outliers could also cause the distributions of differences to not be normal which prevented the use of t-tests, which are statistically stronger tests than the significance tests. Identification and elimination of reasons for extreme outliers in

differences between observer measurements could allow for a stronger statistical test of differences.

Re-examination of the images revealed difficulties experienced by both the novice and the automated system that explain problems in detection and measurements. It was found that the novice consistently missed pores that were not easily distinguishable and was often unable to discern the correct number of pores in pore-dense regions (Figure 6).


**Figure 6. Example of pore dense region and detection issues.** The expert (far left), the automated system (middle), and the novice (far right) often differed in the number of pores detected in a region dense with pores.

When the incorrect number of pores was identified in a region, all true pores were counted as missed and the pores identified were counted as false pores. Therefore, incorrect identification of the number of pores in a region contributed to both lower detection accuracy and a higher false pore identification rate for the novice. While the automated system also had issues correctly identifying pores in pore-dense regions, the main accuracy issue for the automated system was its failure to identify larger sized pores. The primary reasons for the high false pore detection rates by the automated system were determined to be due to a large number of pores identified outside the cortical region, as seen in Figure 7, and identification of groups of osteocytic lacunae located close to one another. Osteocytic lacunae are small holes in which osteocytes reside, but

they have been shown to not significantly influence the mechanical response of bone (Martin, 1984).

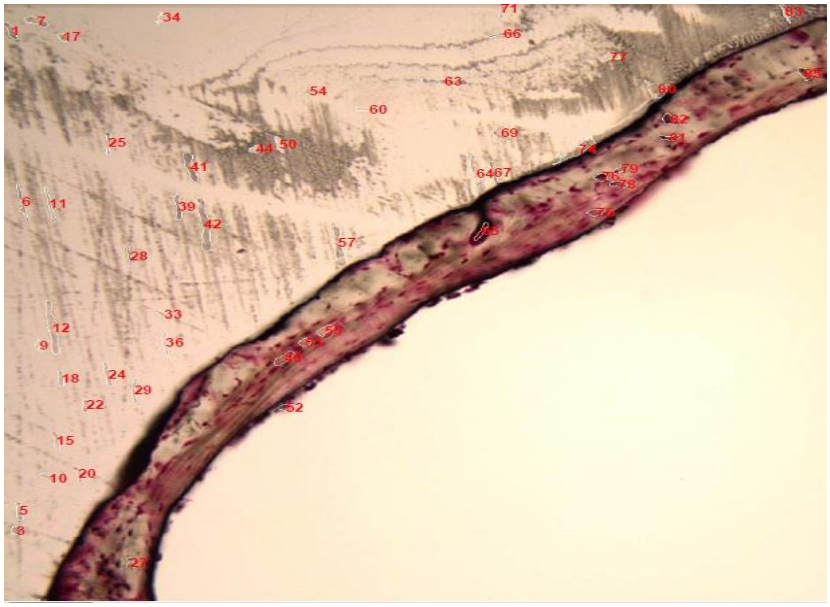


**Figure 7. Automated system detection of pores outside cortex.** Red numbers indicate regions where the automated system detected a pore. High densities of pores are found outside the cortical region, on the background of the slide.

Extreme outliers for measurement differences seemed to be caused by pores with abnormal borders for both the novice and the automated system, as shown in Figure 8. The automated system also exhibited outliers due to pore border detection issues caused by bubbles (Figure 9).


**Figure 8. Example of pore with abnormal borders in detection issues.** The expert (top), automated system (middle) and novice (bottom) often identified different borders for pores that had atypical borders.

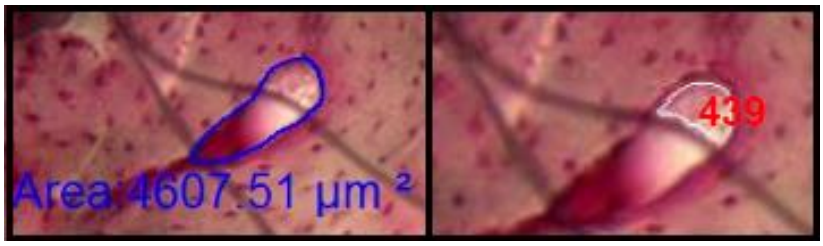


**Figure 9. Example of border detection problem in automated system due to bubbles.** Bubbles on slides often caused the automated detection system to incorrectly define pore borders, resulting in inaccurate area measurements. In this example the automated system (right) defines a smaller area than the expert (left).

While the results of this portion of the analysis conclude that a large-scale study should proceed using an expert for feature identification and measurement, image investigations identified consistent problems with the novice and automated system that could potentially be fixed. It is thought that with minor training sessions, the novice accuracy in identification and measurement of pores could be greatly improved. It is also thought that with minor improvements to the automated system, its accuracy in identification and measurement could be improved and its false pore detection rate could be lowered. Therefore, the novice and automated

system both show promise for future research, but until improvements have been made and evaluated, the expert should be utilized for feature identification and measurement.

### 5.3 Statistical Models

It was hypothesized that model 2 would produce results more similar to the original model than model 1; however, it was found that model 1 exhibited less differences from the original model than model 2. It was found that model 1 differed in results obtained from the original model in cortex variation for percent porosity and crack surface density and in subject variation for crack length. The differences in cortex variation are thought to have been due to the significant differences in the analysis approaches of the original model and model 1. The original model averaged the values for each feature for both of the cutaneous and pleural cortices for each subject and performed a paired t-test. Model 1 did not average any of the values and performed a nested, mixed-model ANOVA. The paired t-test considers only the cortex differences when calculating significant differences, while the ANOVA procedure considers the variation contributions of the subject and rib side in its analysis. It is thought that the loss of information from averaging the values in paired t-tests resulted in the significant difference observed by model 1, but not by the original model for cortex variation in crack surface density. It is also thought that when cortex variation alone is considered in the statistical tests, it appears to be significantly different as in the original model, but it appears not significantly different when considered alongside other sources of variation such as subject or rib side, as in model 1. This thought is supported by the weakly significant p-value of 0.04 in the original model. Lack of significant subject variation for crack length in model 1 was a surprising result since both models

utilized ANOVA procedures to analyze the variables. This could again be due to differences in the variables each model considers when evaluating sources of variance. As few discrepancies in between model 1 and the original model were observed, it is thought that model 1 is a good fit for the data and can be utilized in future analyses. Model 2 exhibited multiple differences from the original model, which were all unexpected and were not able to be explained. It is unlikely that the differences observed were due to differences in the models, since both models used ANOVAs on the same set of variables to analyze variance. The more likely explanation is that model 2 does not fit the dataset accordingly, and therefore cannot accurately assess the variation trends. Additionally, model 2 often produced different significances in the pleural and cutaneous cortices, which makes it difficult to interpret the overall significance of a variable's variation within the entire rib. Due to its unexplainable discrepancies, its inability to assess cortex variation and the confusing nature of its results, model 2 should not be used in future analysis.

Both models were utilized to determine if there was a significant difference in the microfeatures for the inferior and superior sides of the cutaneous and pleural cortices. As hypothesized, there was no significant variation observed when utilizing either model 1 or model 2. This implies that future analyses need not include side of cortex as a variable. This also implies that microfeature measurement can be done in rib halves instead of rib quadrants.

Evaluation of percent porosity in the independent pediatric dataset revealed no significant variation in rib side or rib level and significant cortex variation, which was consistent with the hypothesized results. The observed lack of significant subject variation was an unexpected result; however, it is possible that in pediatric datasets significant subject variation is not present for percent porosity due to the complex and changing material and structural nature of pediatric ribs.

Lack of significant variation in rib level implies that bilateral rib pairs from multiple thoracic levels can be utilized in a large-sample study of bone quality and corresponding biomechanical response. This allows for an increased number of ribs for analysis, as multiple bilateral rib pairs can be obtained from one individual.

### 5.5 Future Directions

This study found that while the novice and the automated system were not yet viable options for replacing the expert in identifying percent porosity, they showed promise with some improvements. Future methodological developments should explore the impact of modifications to the automated system and providing training to the novice in pore identification and measurement accuracies. This study also only explored feature detection method alternatives for percent porosity. Alternative methods for detecting and measuring microfractures should be explored for further improvements in feature detection.

## Chapter 6: Conclusions

The purpose of this study was to improve upon the methodologies introduced in Agnew (2011) to lay the groundwork for the development of a large-sample study of bone quality and biomechanical response. Three new methodological approaches that aid in feasibility of a large-sample analysis of bone microstructure have been defined in this study. The first new approach involved utilizing automated image stitching software programs to facilitate the obtainment of partial or full image reconstructions. The second new approach included utilizing either an automated detection system or trained novices to facilitate microfeature detection and measurement in images of rib cross sections. The final approach defined in this study was the use of a standardized statistical model for large-sample analyses. These three approaches can be built into the framework for an efficient, standardized methodology to investigate the relationship between bone quality and biomechanical response in a normal population.

## References Cited


Agnew, AM. *Histomorphometry of the elderly rib: A methodological approach with implications for biomechanics, function and fracture risk* [dissertation]. Columbus, OH: The Ohio State University; 2011.

Agnew AM, Moorhouse K, Kang YS, et al. The response of pediatric ribs to quasi-static loading: mechanical properties and microstructure. *Annals of Biomedical Engineering*. In press 2013.

Burr DB, Forwood MR, Fyhrie DP, Martin RB, Schaffler MB, Turner CH. Bone microdamage and skeletal fragility in osteoporotic and stress fractures. *Journal of Bone and Mineral Research.* 1997; 12(1):6-15.

Burr DB, Turner CH, Naick P, et al. Does microdamage accumulation affect the mechanical properties of bone? *Journal of Biomechanics* 1998; 31(4):337-345.

Ebacher V, Tang C, McKay H, Oxland TR, Guy P, Wang R. Strain redistribution and cracking behavior of human bone during bending. *Bone* 2007; 40(5):1265-1275.

Frost HM. Emerging views about "osteoporosis", bone health, strength, fragility, and their determinants. *Journal of Bone and Mineral Metabolism* 2002; 20(6):319-325.

Laval-Jeantet A, Bergot C, Carroll R, Garcia-Schaefer F. Cortical Bone Senescence and Mineral Bone Density of the Humerus. *Calcified Tissue International* 1983; 35(1):268-272.

Martin RB. Porosity and specific surface of bone. *CRC Critical Reviews in Biomedical Engineering*. 1984; 10(3):179-222.

Rose, DC. *The use of geographical information systems software for the spatial analysis of bone microstructure* [dissertation]. Columbus, OH: The Ohio State University; 2011.

Sobelman OS, Gibeling JC, Stover SM, et al. Do microcracks decrease or increase fatigue resistance in cortical bone? *Journal of Biomechanics* 2004; 37(9):1295-1303.

Taylor D. Bone maintenance and remodeling: a control system based on fatigue damage. *Journal of Orthopaedic Research* 1997; 15:601-606.

Turner, CH. Biomechanics of bone: Determinants of skeletal fragility and bone quality. *Osteoporosis International* 2002; 3:97-104.

Zioupos P. Ageing human bone: Factors affecting its biomechanical properties and the role of collagen. *Journal of Biomaterials Applications* 2001; 15(3):187-229.